Introduction to Hypothesis Testing

VT-DSPG 2025

Anbin Rhee

June 27, 2025

Contents

Why Do We Need Hypothesis Testing?

As a data analyst, we often ask questions about data

- Does Ad Campaign A perform better than Ad Campaign B?
- Did the new teaching method improve student scores?
- Is the average height of students in our class the same as the national average?

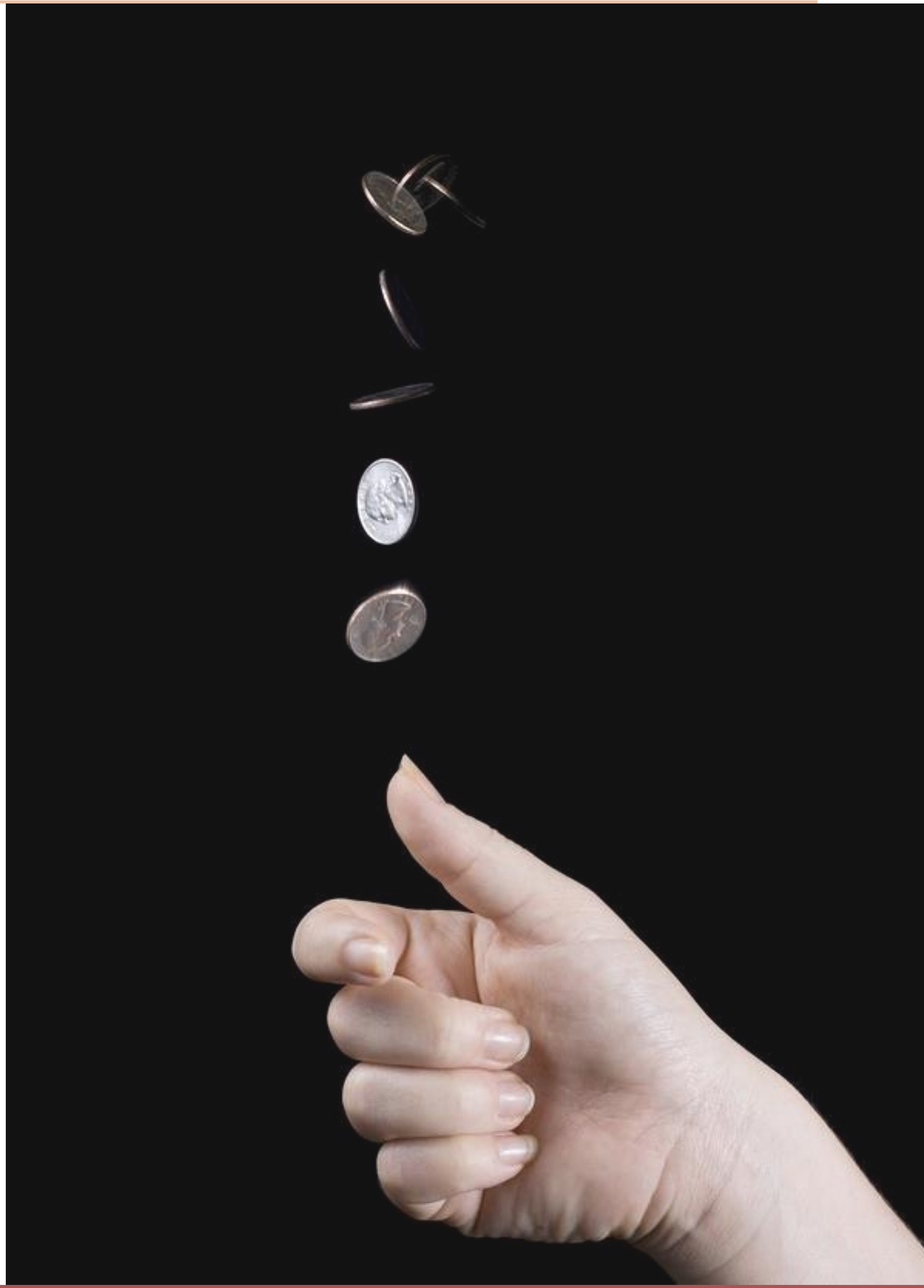
Why Do We Need Hypothesis Testing?

Data can give us clues, but are these observations **real** or just due to **chance**?

Example: If you flip a coin 10 times and get 9 heads, is the coin really unfair, or did you just get lucky?

We need a formal rule so decisions are not based on gut feeling

Hypothesis testing is a **statistical tool** that helps us answer this question!



What is a Hypothesis?

- A hypothesis is a testable statement or educated guess about a population or phenomenon.
- In hypothesis testing, we usually set up two competing hypotheses:
  a. The status quo / no effect hypothesis
  b. The alternative / effect hypothesis

Null Hypothesis (H_0) : The Status Quo

- The ‘default’ assumptions. It states there is **no effect, no difference,** or **no relationship**.
- Think of it like being “innocent until proven guilty” in a courtroom. We assume H_0 is true unless we find strong evidence against it.

Null Hypothesis (H_0) : The Status Quo

- Examples:
  - H_0 : Ad Campaign A and Ad Campaign B have no difference in effectiveness
  - H_0 : The new teaching method has no effect on student scores
  - H_0 : The average height of students in our class is the same as the national average

Alternative Hypothesis (H_1) : Our Claim

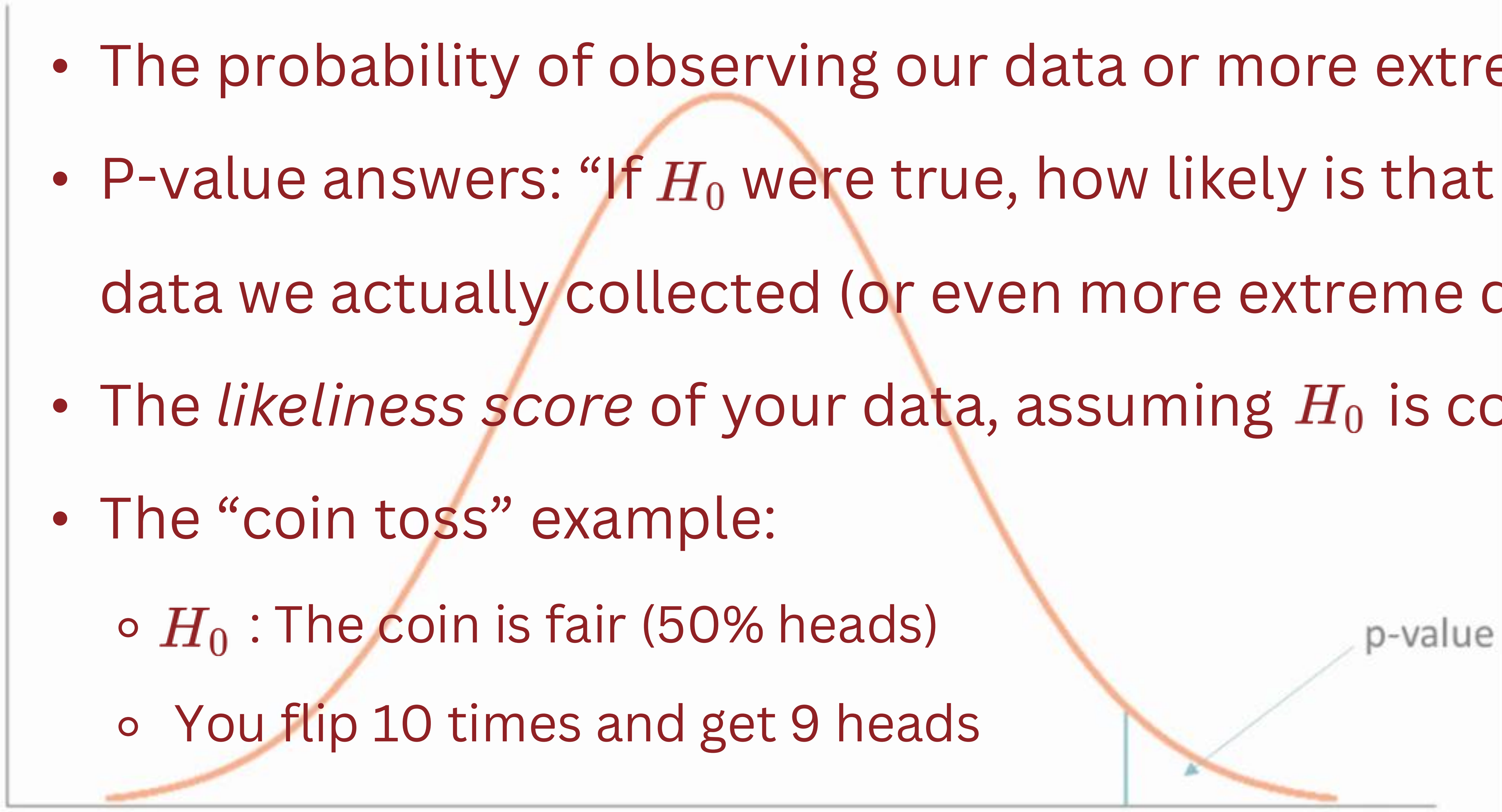
- The hypothesis we are trying to find evidence for.
- It contradicts the null hypothesis.
- It states there is **an effect**, **a difference**, or **a relationship**.

Alternative Hypothesis (H_1) : Our Claim

- Examples:
    - H_1 : Ad Campaign A is more effective than Ad Campaign B
    - H_1 : The new teaching method improves student scores
    - H_1 : The average height of students in our class is different from the national average

P-value

- The probability of observing our data or more extreme data by chance.
- P-value answers: “If H_0 were true, how likely is that we would observe the data we actually collected (or even more extreme data) purely by chance?”
- The *likeliness score* of your data, assuming H_0 is correct
- The “coin toss” example:
  - H_0 : The coin is fair (50% heads)
  - You flip 10 times and get 9 heads



P-value: What is the probability of getting 9 or more heads in 10 flips if the coin is truly fair?

Interpreting the P-value

- It is a metric to answer "Is our evidence strong enough?"
- Small p-value (e.g. 0.0001):
  - If H_0 were true, seeing our data would be very rare
  - Conclusion: We have strong evidence against H_0. We reject the null hypothesis
- Large p-value (e.g. 0.8):
  - If H_0 were true, seeing our data would be quite common
  - Conclusion: We fail to reject the null hypothesis

The Significance Level (α):

- How ‘small’ is small enough for a p-value?
- We define a threshold called the **significance level** (α)
- Common α values: 0.05 (5%) or 0.01(1%)
- Decision Rule:
  - If p-value $<$ α : Reject H_0 (The results is statistically significant)
  - If p-value $\geq$ α : Fail to reject H_0 (The result is not statistically significant)

No Guarantee in Statistics

- Even with statistical test, we are dealing with **probabilities**, not certainties.
- There is always a chance we make the wrong decision based on our sample data

Type I Error: False Positive

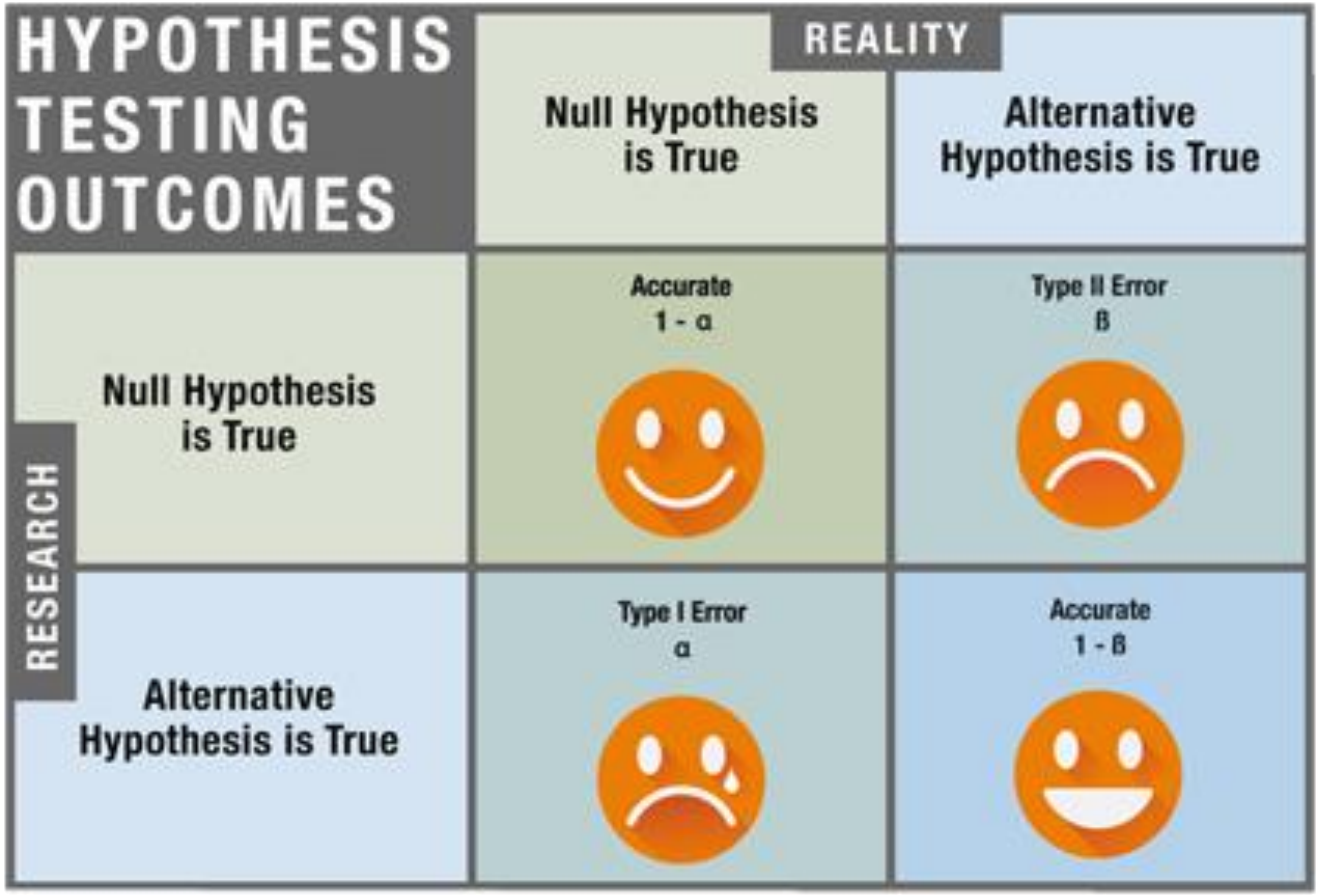
VIRGINIA TECH

- Definition: We reject the Null Hypothesis, but in reality, H_0 was actually true.
- The Risk: The probability of committing a Type I Error is equal to our significance level, α
  - If $\alpha = 0.05$, there is a 5% chance of making a Type I error
- Example: Concluding a person is guilty when in reality the person is innocent

Type II Error: False Negative

- Definition: We fail to reject the Null Hypothesis, but in reality, H_0 was actually false.
- The Risk: The probability of committing a Type II Error is denoted as β which is related to statistical power
  - Statistical power = 1 - β
- Example: A person is truly guilty, but we found them not guilty due to insufficient evidence

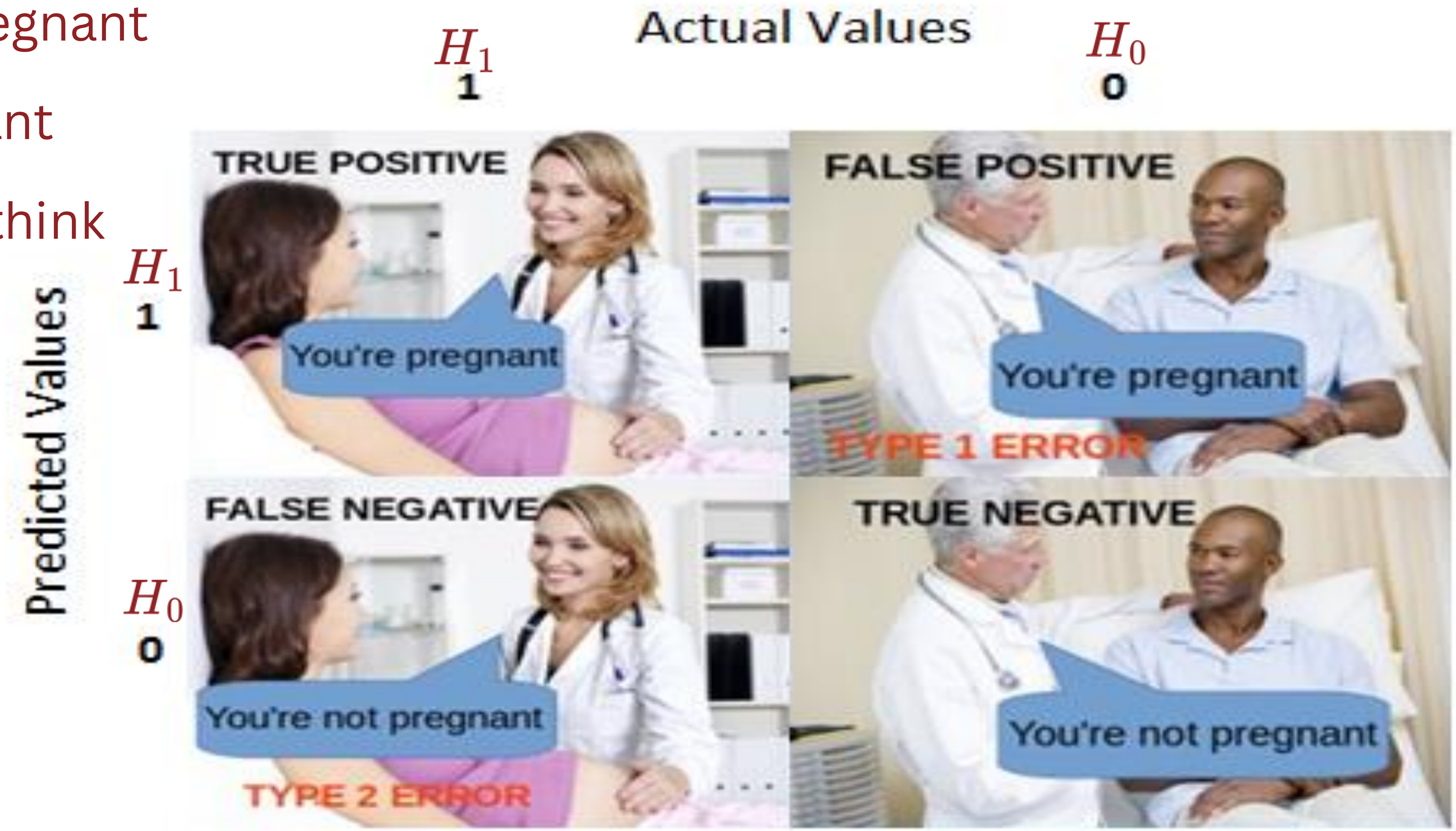
Trade-off between Type I and Type II Errors



- Trade-off: Lowering α reduces Type I error, but raises Type II error.
- The "best" α depends on the consequences of each error in your specific situation.

Trade-off between Type I and Type II Errors

- H_0: Patient is not pregnant
- H_1: Patient is pregnant
- Which error do you think is more costly or dangerous?



The Hypothesis Testing Process

1. Formulate Your Question
2. State Your Hypotheses (H_0 & H_1)
3. Collect Data and Perform Statistical Test
4. Calculate the p-value
5. Make a Decision : Compare p-value to α (Reject H_0 or not?)
6. Interpret Your Conclusion

A Quick Example

VIRGINIA TECH

A Company develops a new fertilizer and wants to know if it increases crop yield

- Step 1: Does the new fertilizer increase crop yield?
- Step 2: Define hypotheses
    - H_0 : The new fertilizer has no effect on crop yield
    - H_1 : The new fertilizer increases crop yield
- Step 3: Imagine we run an experiment and perform a statistical test
- Step 4: Our analysis gives a p-value = 0.02 (Say our α = 0.05)
- Step 5: Reject H_0!
- Step 6: We have statistically significant evidence to conclude that the new fertilizer increase crop yield. The company can be more confident about new fertilizer

Why Hypothesis Testing Thinking Matters In Data Science

Suppose a company wants to predict which customers are likely to stop using their mobile app

- Modern data science often focuses on prediction, not inference
- Prediction-focused workflow may accurately predict who leave, but it tells us nothing about **why** they leave
- Many real-world questions still require hypothesis testing / inference
- Hypothesis testing provides a framework for rigorous reasoning

Summary

- Hypothesis testing helps use make data-driven decision by assessing evidence
- Understand H_0 and H_1
- The p-value is the probability of observing our data by chance, if H_0 is true
- We reject H_0 when p-value is less than our threshold alpha
- Be aware of Type I error and Type II error
- Integrate hypothesis testing thinking beyond just algorithms to ensure rigor, interpretability, and trustworthiness in your research

Next

- Commonly used statistical models
- When each model might be a good choice
- If there is something you'd like to learn in the next workshop, feel free to let me know!